Interdependencies

Temurshoev, U.

Chapter 7

Epilogue

7.1 Introduction

This thesis has focused on three types of interdependencies: at the levels of firms, individuals and economic sectors. These were firms' shareholding interlocks, social networks of people, and production linkages of industries. Given that these interrelationships have their own distinguishing features, it is obvious that the analytical frameworks used in their analysis were (quite) different as well. This explains why this study did not focus on one specific field, but instead investigated topics from several subfields of economics and sociology, such as Finance, Industrial Organization, Input-Output Economics, Network Economics, and Social Network Analysis. However, on the other hand, the issues considered in this thesis are not at all independent of each other, in contrast to what might appear at first glance. The analyses of the complex webs of interrelations have a lot in common. In some sense they adopt a unified analytical framework, thus extending the frontiers of common interests in the above-mentioned fields.

Some of the main questions in this study aimed at the following. How to quantify ownership complexity caused by cross ownership by both individuals and companies? What is the appropriate measure of separation of ownership and control rights due to firms' cross-shareholdings? Does a firm with passive stockholdings in its rivals exert strictly higher market power than a firm without any shareholdings? Do interfirm shareholding interlocks matter in the empirical study of market performance? What is the effect of partial cross ownership on the incentives of asymmetric (in terms of costs) firms to collude and on the collusive price? How to find the group of individuals with the maximum impact on the overall equilibrium outcome in (social) networks? How to incorporate individuals' exogenous heterogeneity into the analysis of key players search? Is the key sector problem equivalent to the problem of identification of key group of sectors? If no, what are the underlying reasons?

These questions were thoroughly addressed in the previous chapters. In the next section we give a brief summary of the obtained results. The last section discusses three directions for future research, each based on the findings in this thesis.

7.2 Summary of results

Chapter 2 proposed new measures of network complexity due to the existence of cross ownership links among firms. The measures called "weighted average distance of indirect linkages" (WADIL) and "weighted average distance of total linkages" (WADTL) quantify the complexity of an ownership structure that is characterized by crossholdings of stocks. The proposed measures consider both the sizes of direct and indirect shareholdings, and the average distance between the owners and the owned firms. We say that owner (or firm) i has an *indirect* stake in firm r if it has a stake in a firm that has a stake in firm r, or if it has a stake in a firm that has a stake in a firm that has a stake in firm r, and so on. The average distance was obtained from the average number of intermediate firms via whom the ownership link between i and r runs. The values of WADILs and WADTLs indicate whether a certain link is of a direct nature only or whether indirect shareholdings also play a role in the link. The larger values of WADILs and WADTLs indicate a more complex network involving a larger number of different ownership paths. Combining the linkage size and the distance allowed us to visualize the cross-shareholding interlocks and the true ownership relations. The methodology was applied to the Czech banking sector in 1997. It was found that there is ample evidence that indirect ownership relations play a crucial role in the banking sector in the Czech Republic. Further, the link between the proposed measures of network complexity and the degree of separation of dividend and control rights due to cross-shareholdings was explored. It was suggested that the WADILs and the WADTLs may serve as alternative measures for the degree of separation. That is, the more complex the network of non-negligible ownership relations is, the larger is the degree of control enhancement due to cross-shareholding links among firms. As a consequence, also the gap between the control and ownership stakes of owners in firms is larger. This was confirmed by the empirical results for the Czech banking sector. The obtained

WADILs and WADTLs were also compared to the wedges between ownership and control rights, where the last were quantified by well-known methodologies from finance, namely the "weakest link" and the "dominant shareholder" approach.

The effect of disregarding partial cross ownership (PCO) (i.e., shares that do not give control power to their owners) in empirical studies of market performance and firms' market power was investigated in Chapter 3.[1] For this purpose the well-known framework of the "structure-conduct-performance school" in industrial organization was used. For the estimation of firms' market power and the tacit collusion that is inherent to an industry, the framework was modified by including both direct and indirect shareholdings. It was proved that, unlike in the no-PCO case, the link between firms' price-cost margins and the degree of market competitiveness is nonlinear in the presence of PCO. Thus, ignoring PCO in an analysis of an industry with extensive shareholdings between firms, will most likely lead to biased results due to model misspecification. In an empirical application, it was found that Japanese commercial banks in 2003 were competing in a modest collusive environment. However, if PCO was disregarded, the results were different and indicated a Cournot oligopoly. It was further found that banks with PCO in their rivals exert a strictly larger market power than those without any shareholdings. In particular, city banks with many shareholdings were found to exercise a much larger market power than regional banks with none or few stockholdings. Hence, the hypothesis that acquiring shares in rivals for a firm is one of the means of enhancing its market power was confirmed in Chapter 3.

Chapter 4 adopted an infinitely repeated Bertrand oligopoly model to investigate the effect of partial ownership arrangements of firms under cost asymmetries on their incentives to collude. We first considered the case where only the most efficient firm in the industry invests in rivals. It was shown that a unilateral partial ownership by this firm may facilitate a market-sharing scheme in which all firms charge the same collusive price and divide the market between them. We showed that when the most efficient firm invests in rivals, the collusive price, which is a compromise between the monopoly prices of the different firms, increases relative to the case where there are no partial ownership arrangements. Further, we focused on the effect of a change in the PCO structure on tacit collusion. It was shown that when the stake that firm r has in firm s increases at the expense of outside shareholders collusion is never hindered. It will even be strictly facilitated if and only

[1] In Chapter 4 we also consider the case when only one firms invests in rivals. We call this a partial ownership (PO) case. PCO, instead, reflects the fact that in the presence of multilateral ownership arrangements, cross-shareholdings by firms are possible.

if (i) the industry maverick (the firm with the strongest incentive to deviate from a collusive agreement) has a direct or indirect stake in firm r, and (ii) firm s is not the industry maverick. When (i) and/or (ii) fail to hold, the increase in firm r's stake in firm s does not affect tacit collusion. These results extend the earlier findings in Gilo et al. (2006) and show that the results for firms with symmetric cost functions generalize to the asymmetric costs case. Chapter 4 also considered the case of such an ownership change due to transfer of ownership between firms. It was shown that a transfer of partial cross ownership in firm s from firm k to firm r does not affect tacit collusion if the industry maverick is firm s or if, at the outset, the industry maverick has the same total (direct and indirect) share in firms k and r. Otherwise, the transfer of partial cross ownership facilitates tacit collusion if the industry maverick has a larger total share in firm r (the acquirer) than in firm k (the seller) but hinders tacit collusion if the reverse holds.

Chapter 5 extended the problem of finding the key player in a network game studied by Ballester et al. (2006) to the search of the key group, where players' exogenous heterogeneity was taken into account. The key group is the group of players that has the maximum (or minimum) possible impact on the overall equilibrium activity level of the network. We derived a closed-form expression of the so-called group intercentrality measure, which is used to identify the key group in networks, and explored some of its properties. Further, the measures of weighted and unweighted group intercentralities that depend only on the initial network configuration were shown to be useful for the identification of the key group of heterogeneous players. The weights are based on observable differences of players, such as age, education, occupation, race, religion, family size, or parents' education. It was shown that once these observable differences are accounted for, the results of the key player/group problem may significantly change when compared to the results based on the assumption of homogeneous players. Finally, the size of the key group was endogenized, which is an important issue since targeting groups of different sizes incurs different benefits and costs. Hence, from the planner's perspective it is essential to get an idea of what is the optimal size of the key group, i.e., what size yields the largest net benefit.

Chapter 6 investigated the issue of finding key sectors of an economy, that is, sectors with the maximum potential of spreading growth impulses throughout the economy and thus impacting output or some other factor (such as value added, employment, or CO_2 emissions). For this purpose, the hypothetical extraction method (HEM) from input-output analysis was adopted, which measures the contribution of each sector to the overall gross output or any other factor by comparing the original result with the result that is obtained from omitting one sector (or a group of sectors) from the model. The reduction in, for example, output is due to this omission and thus reflects the role of the hypothetically extracted (group of) sector(s). Explicit formulations of the optimization problems of finding a key sector and a key group of sectors from the HEM perspective were given, and their analytical solutions (called industries' factor worths) were derived. It was shown that the key group of $k \geq 2$ sectors is, in general, different from the k sectors with the largest individual contributions to the overall factor production/consumption, which was confirmed in an example of the Australian economy in case of water use and CO_2 emissions in the mid of 1990s. This outcome has to do with the fact that in reality sectors may be redundant with respect to each other if they have similar patterns and sizes of production linkages, final demands and factor production capabilities. The related issues of finding a key region and key group of regions in an interregional input-output (IO) framework were investigated similarly. Further, we showed that the factor worth measure is invariant to the netting out of intrasectoral transactions for any factor other than gross output. Hence, the outcomes of the key sector/group problems in the standard and the so-called net input-output settings are exactly identical so long as the factor is not total output. The link of the HEM problems to the fields of influence approach was pointed out, which gives an alternative economic interpretation of these problems in terms of the economy-wide effects of an incremental change in sectors' input self-dependencies. Finally, it was proved that an increase (decrease) in an input coefficient never decreases (increases) the factor worth/importance of any sector, and the conditions for a subsequent strict change were derived.

7.3 Related future research

Often, doing research raises new issues. In this respect I absolutely agree with my supervisor Erik Dietzenbacher, who in the final chapter of his dissertation states: "Answers raise new questions, solutions define new problems, results call for a generalization or a sharpening, assumptions for a relaxation, gaps need to be filled up, and loose ends are to be tied up" (Dietzenbacher, 1991, p. 267). Hence, in what follows I will present three directions for future research, the basis of which is essentially the current study.

7.3.1 Engines of growth: a hypothetical extraction approach

In Chapter 6, the problem of the identification of the key sectors for generating some economic, social, and/or environmental factor was discussed. A similar approach can be applied to the identification of "key sectors" for generating economy-wide total factor productivity (TFP) growth. In the literature, such sectors are called the *engines of growth*. However, the generalized hypothetical extraction method (HEM) as discussed thoroughly in Chapter 6 is not adequate to deal with finding the engines of growth, because TFP growth cannot be directly incorporated into the input-output framework and needs a somewhat different setting. Such a framework will be discussed below after we briefly present the analysis of productivity spillovers.

7.3.1.1 Productivity analysis of spillovers

Ten Raa and Wolff (2000) propose to identify the engines of growth as follows. The departing point is the Solow (1957) residual definition of total factor productivity (TFP) growth, g:

$$g = \frac{\mathbf{p}'d\mathbf{f} - wdL - rdK}{\mathbf{p}'\mathbf{f}}, \tag{7.1}$$

where $\mathbf{f}$ is the final demand vector (also termed net output vector), L and K are, respectively, labor and capital inputs, w and r are their respective prices, and $\mathbf{p}$ is the vector of production prices. These prices reflect zero profits since

$$\mathbf{p}'(\mathbf{I} - \mathbf{A}) = \mathbf{v}' = w\mathbf{l}' + r\mathbf{k}', \tag{7.2}$$

where $\mathbf{A}$ is the input matrix, and $\mathbf{v}$, $\mathbf{l}$ and $\mathbf{k}$ are the vectors of direct value-added, labor and capital coefficients.

Using the balancing equation of the open Leontief model $\mathbf{f} = (\mathbf{I} - \mathbf{A})\mathbf{x}$, where $\mathbf{x}$ is the vector of total (or gross) outputs, the numerator of (7.1) can be written as

$$\begin{aligned} \mathbf{p}'d\mathbf{f} - wdL - rdL &= \mathbf{p}'d(\mathbf{I} - \mathbf{A})\mathbf{x} - wd(\mathbf{l}'\mathbf{x}) - rd(\mathbf{k}'\mathbf{x}) \\ &= (-\mathbf{p}'d\mathbf{A} - wd\mathbf{l}' - rd\mathbf{k}')\mathbf{x} + (\mathbf{p}'(\mathbf{I} - \mathbf{A}) - w\mathbf{l}' - r\mathbf{k}')d\mathbf{x}, \end{aligned} \tag{7.3}$$

where the last term vanishes if we use the production prices in (7.2). Using (7.3), (7.1) reduces to

$$g = \frac{-(\mathbf{p}'d\mathbf{A} + wd\mathbf{l}' + rd\mathbf{k}')\mathbf{x}}{\mathbf{p}'\mathbf{f}} = \frac{\boldsymbol{\pi}'\hat{\mathbf{p}}\mathbf{x}}{\mathbf{p}'\mathbf{f}}, \tag{7.4}$$

where $\boldsymbol{\pi}' = -(\mathbf{p}'d\mathbf{A} + wd\mathbf{l}' + rd\mathbf{k}')\hat{\mathbf{p}}^{-1}$ is the row vector of sectoral TFP growth rates, and $\hat{\mathbf{p}}\mathbf{x}/(\mathbf{p}'\mathbf{f})$ is the vector of so-called Domar weights.

Spillovers are measured as a weighted average of the TFP growth in supplying sectors. Four explanatory variables for the TFP growth rate of sector j, π_j, are distinguished: (1) an autonomous source, α, (2) R&D in sector j per dollar of gross output, $\rho = RD_j/(p_j x_j)$, (3) a direct productivity spillover, $\sum_i (p_i a_{ij}/p_j)\pi_i$, and (4) a capital embodied spillover, $\sum_i (p_i b_{ij}/p_j)\pi_i$, where b_{ij} is the capital stock coefficient of capital good i in sector j. This yields the following regression equation (Wolff, 1997):

$$\boldsymbol{\pi}' = \alpha\boldsymbol{\imath}' + \beta_1\boldsymbol{\rho}' + \beta_2\boldsymbol{\pi}'\hat{\mathbf{p}}\mathbf{A}\hat{\mathbf{p}}^{-1} + \beta_3\boldsymbol{\pi}'\hat{\mathbf{p}}\mathbf{B}\hat{\mathbf{p}}^{-1} + \boldsymbol{\varepsilon}', \tag{7.5}$$

where $\boldsymbol{\varepsilon}$ is the vector of error terms.

Let us denote the spillover matrix by

$$\mathbf{C} \equiv \beta_2\hat{\mathbf{p}}\mathbf{A}\hat{\mathbf{p}}^{-1} + \beta_3\hat{\mathbf{p}}\mathbf{B}\hat{\mathbf{p}}^{-1} = \hat{\mathbf{p}}\left[\beta_2\mathbf{A} + \beta_3\mathbf{B}\right]\hat{\mathbf{p}}^{-1}, \tag{7.6}$$

then, ignoring the error term, (7.5) can be rewritten as

$$\boldsymbol{\pi}' = \alpha\boldsymbol{\imath}' + \beta_1\boldsymbol{\rho}' + \boldsymbol{\pi}'\mathbf{C}. \tag{7.7}$$

Plugging (7.7) back in (7.4) yields

$$g = \frac{[\alpha\boldsymbol{\imath}' + \beta_1\boldsymbol{\rho}' + \boldsymbol{\pi}'\mathbf{C}]\hat{\mathbf{p}}\mathbf{x}}{\mathbf{p}'\mathbf{f}} = \alpha DR + \beta_1\frac{RD'\boldsymbol{\imath}}{\mathbf{p}'\mathbf{f}} + \frac{\boldsymbol{\pi}'\mathbf{C}\hat{\mathbf{p}}\mathbf{x}}{\mathbf{p}'\mathbf{f}}, \tag{7.8}$$

where RD is the vector of sectoral R&Ds, $\boldsymbol{\imath}$ is the summation vector, and $DR = \mathbf{p}'\mathbf{x}/\mathbf{p}'\mathbf{f}$ is the Domar ratio.

Equation (7.8) gives the *direct* effect of R&D on TFP growth. β_1 measures the *direct rate of return to R&D intensity* or, equivalently, the *direct return to R&D*, in terms of output value per dollar expenditure, because the denominator in the definition g in (7.1) is also the dollar value of expenditures, i.e., $\mathbf{p}'\mathbf{f}$.

The *total* returns to R&D, however, are obtained by taking into account the spillover effects, captured by the last term in (7.8). Define $\mathbf{M} \equiv (\mathbf{I} - \mathbf{C})^{-1} =$

$\mathbf{I} + \mathbf{C} + \mathbf{C}^2 + \cdots$. Solving (7.7) for the vector of TFP growth rates gives

$$\boldsymbol{\pi}' = (\alpha \boldsymbol{\imath}' + \beta_1 \boldsymbol{\rho}')\mathbf{M}, \tag{7.9}$$

hence, using (7.4)

$$g = \frac{\boldsymbol{\pi}'\hat{\mathbf{p}}\mathbf{x}}{\mathbf{p}'\mathbf{f}} = \frac{(\alpha \boldsymbol{\imath}' + \beta_1 \boldsymbol{\rho}')\mathbf{M}\hat{\mathbf{p}}\mathbf{x}}{\mathbf{p}'\mathbf{f}}. \tag{7.10}$$

Therefore the *total rate of return to R&D intensity* ρ_i amounts to $\beta_1(\sum_j m_{ij} p_j x_j)/\mathbf{p}'\mathbf{f}$. Here, β_1 is inflated by multipliers m_{ij} because of spillover effects and also by gross/net output ratios as the sectoral R&D intensities ρ_i are defined as the R&D/gross output ratios. Note that while the first decomposition in (7.4) is a TFP growth accounting identity, the second decomposition in (7.10) attributes TFP growth to sources of growth taking into account the spillover effects.

Since $\boldsymbol{\rho}' = RD'(\hat{\mathbf{p}}\hat{\mathbf{x}})^{-1}$, we have that

$$g = \alpha \frac{\boldsymbol{\imath}'\mathbf{M}\hat{\mathbf{p}}\mathbf{x}}{\mathbf{p}'\mathbf{f}} + \beta_1 \frac{RD'(\hat{\mathbf{p}}\hat{\mathbf{x}})^{-1}\mathbf{M}\hat{\mathbf{p}}\mathbf{x}}{\mathbf{p}'\mathbf{f}},$$

hence the *total return to R&D*, in terms of output value per dollar expenditure in sector i, amounts to $\beta_1(\sum_j m_{ij} p_j x_j)/(p_i x_i)$. So the direct return to β_1 is inflated by the factor $(\sum_j m_{ij} p_j x_j)/(p_i x_i)$ because of spillover effects stemming from sector i. Since the factors $(\sum_j m_{ij} p_j x_j)/(p_i x_i)$ reinforce the returns to R&D, they are *spillover multipliers*. Hence, the vector of spillover multipliers is given by $(\hat{\mathbf{p}}\hat{\mathbf{x}})^{-1}\mathbf{M}\hat{\mathbf{p}}\mathbf{x}$. Spillover multipliers are equal to the ratio of the total to the direct return to R&D, thus measure the external effects of sectoral R&D.

From (7.10) it follows that the overall TFP growth g is decomposed into sources of growth $\alpha + \beta_1 \rho_i$ aggregated by the linkages $\sum_j m_{ij} p_j x_j$ for sector $i = 1, \ldots, n$. Sectors that contribute much to overall TFP growth in this decomposition are the *engines of growth*. For example, the largest engine of growth is the sector, say, i, with the largest value of $(\alpha + \beta_1 \rho_i) \sum_j m_{ij} p_j x_j$.

7.3.1.2 Engines of growth from a hypothetical extraction perspective

For simplicity, let us denote the vector of *spillover linkages* by $\mathbf{s} \equiv \mathbf{M}\hat{\mathbf{p}}\mathbf{x}$. The sectoral direct and indirect productivity gains are thus given by the vector $(\alpha \mathbf{I} + \beta_1 \hat{\boldsymbol{\rho}})\mathbf{s}$ as derived in the previous section. ten Raa and Wolff (2000) define the engines of growth as the 10 sectors with the largest values in the last vector.

Let us now consider the problem of identifying the engines of growth from the HEM approach, i.e., sectors whose elimination from the systems of economy-wide production and spillover interrelations causes the largest reduction in the overall TFP growth rate. Consider the identification of $k \in [1, n-1]$ engines of growth. Denote by $\mathbf{C}^{-\{i_1,\ldots,i_k\}}$ the new spillover matrix derived from $\mathbf{C}$ by setting to zero all its i_s-th rows and columns elements, where $s = 1, \ldots, k$. From (7.6), it follows that this is equivalent to nullifying all rows and columns entries corresponding to $i_1, \ldots, i_k$ of the input matrix $\mathbf{A}$ and the capital stock coefficient matrix $\mathbf{B}$. The assumption therefore is that in the new system without sectors $i_1, \ldots, i_k$ the production and capital stock structures of other active sectors $j \notin \{i_1, \ldots, i_k\}$ remain unchanged.[2] Although at first glance this assumption seems restrictive, in fact it is not, given our main aim of identifying the importance of sectors $i_1, \ldots, i_k$ in generating nation-wide growth considering the spillover effects. The point is that by taking all other input and capital stock coefficients fixed, we explicitly allow the resulting outcome to depend only on the extraction of sectors $i_1, \ldots, i_k$, which are now not participating in the "roundabout" of the production process, hence not contributing to the TFP growth either. The vector of spillover linkages after extracting sectors $i_1, \ldots, i_k$ is $\mathbf{s}^{-\{i_1,\ldots,i_k\}} = \mathbf{M}^{-\{i_1,\ldots,i_k\}}\hat{\mathbf{p}}\mathbf{x}^{-\{i_1,\ldots,i_k\}}$, where $\mathbf{M}^{-\{i_1,\ldots,i_k\}} = (\mathbf{I} - \mathbf{C}^{-\{i_1,\ldots,i_k\}})^{-1}$, and $\mathbf{x}^{-\{i_1,\ldots,i_k\}} = (\mathbf{I} - \mathbf{A}^{-\{i_1,\ldots,i_k\}})^{-1}\mathbf{f}^{-\{i_1,\ldots,i_k\}}$. The new net output vector $\mathbf{f}^{-\{i_1,\ldots,i_k\}}$ is the same as $\mathbf{f}$ except its i_1-th, ..., i_k-th entries that are all set to zero. The reason for setting $f_{i_s} = 0$ for all $s = 1, \ldots, k$ is that when sectors $i_1, \ldots, i_k$ cease to exist, their (domestic) gross outputs should be zero.

We further denote the sum of autonomous source and R&D intensities by $\boldsymbol{\lambda} \equiv \alpha\boldsymbol{\imath} + \beta_1\boldsymbol{\rho}$. Given the vectors of sources of growth $\boldsymbol{\lambda}$ and production prices $\mathbf{p}$, the objective is picking k ($1 \leq k \leq n-1$) sectors $i_1, i_2, \ldots, i_k$ ($i_s \neq i_r$) such that their extraction from the economy generates the highest possible reduction in the overall TFP growth rate, $g = \boldsymbol{\lambda}'\mathbf{s}/\mathbf{p}'\mathbf{f}$. Formally, the problem is

$$\max\left\{\frac{\boldsymbol{\lambda}'\mathbf{s}}{\mathbf{p}'\mathbf{f}} - \frac{\boldsymbol{\lambda}'\mathbf{s}^{-\{i_1,\ldots,i_k\}}}{\mathbf{p}'\mathbf{f}^{-\{i_1,\ldots,i_k\}}} \;\middle|\; \{i_1,\ldots,i_k\} \subseteq \{1,\ldots,n\}; i_s \neq i_r\right\}. \tag{7.11}$$

This is a finite optimization problem, which admits at least one solution. The solution to (7.11) is denoted by $\{i_1^*, i_2^*, \ldots, i_k^*\}$ and is called the *k engines of growth*. Removing these industries from the initial input and capital stock structures have the largest impact on the overall TFP growth rate.

[2] This is usual for all the HEM approaches, the only difference now is that besides production we also consider the capital stock structure.

Before giving an explicit solution to the problem (7.11), first, recall that $\mathbf{f} = \mathbf{x} - \mathbf{Ax}$. Its premultiplication by the diagonal matrix of the (fixed) production prices yields $\hat{\mathbf{p}}\mathbf{f} = \hat{\mathbf{p}}\mathbf{x} - \hat{\mathbf{p}}\mathbf{A}\hat{\mathbf{p}}^{-1}\hat{\mathbf{p}}\mathbf{x}$, hence $\hat{\mathbf{p}}\mathbf{x} = (\mathbf{I} - \hat{\mathbf{p}}\mathbf{A}\hat{\mathbf{p}}^{-1})^{-1}\hat{\mathbf{p}}\mathbf{f}$. Thus, in what follows in this section the Leontief inverse is defined as $\mathbf{L} \equiv (\mathbf{I} - \hat{\mathbf{p}}\mathbf{A}\hat{\mathbf{p}}^{-1})^{-1}$.

Let $\mathbf{E}$ be the $n \times k$ matrix defined as $\mathbf{E} = \left(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k}\right)$, where $\mathbf{e}_i$ is the i-th column of the identity matrix. Each column of $\mathbf{E}$ has all zeros except one positive number being unity that corresponds to one of the extracted sectors. Hence, the reduced multiplier matrix $\mathbf{M}_{kk} = \mathbf{E}'\mathbf{ME}$ includes all the elements of the original multiplier matrix $\mathbf{M}$ that are directly related to the extracted sectors $i_1, \ldots, i_k$. Similarly, the reduced Leontief inverse is $\mathbf{L}_{kk} = \mathbf{E}'\mathbf{LE}$. Recall from (7.9) that the vector of sectoral TFP growth rates is $\boldsymbol{\pi}' = \boldsymbol{\lambda}'\mathbf{M}$. Next note that the problem in (7.11) is equivalent to

$$\min \left\{ \frac{\boldsymbol{\lambda}'\mathbf{s}^{-\{i_1,\ldots,i_k\}}}{\mathbf{p}'\mathbf{f}^{-\{i_1,\ldots,i_k\}}} \;\middle|\; \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}; i_s \neq i_r \right\}.$$

The "residual" TFP growth rate in the objective above is defined as the *reduced TFP growth* due to extraction of sectors $i_1, \ldots, i_k$ ($i_r \neq i_s$), and it can be shown to be equal to (see Appendix 7.A)

$$g^r_{i_1,\ldots,i_k} = \frac{\boldsymbol{\pi}'\left(\mathbf{I} - \mathbf{EM}_{kk}^{-1}\mathbf{E}'\mathbf{M}\right)\left(\mathbf{I} - \mathbf{LEL}_{kk}^{-1}\mathbf{E}'\right)\hat{\mathbf{p}}\mathbf{x}}{\mathbf{p}'(\mathbf{I} - \mathbf{EE}')\mathbf{f}}. \tag{7.12}$$

Notice that, disregarding the denominator in (7.12), when $k = n$ and $\mathbf{E} = \mathbf{I}$ the numerator of the reduced TFP growth due to extraction of *all* sectors in (7.12) becomes zero, which is entirely expectable because without an industry, the hypothetical total (direct and indirect) productivity gains should be zero, i.e., $\boldsymbol{\lambda}'\mathbf{s}^{-\{1,\ldots,n\}} = 0$. We have established the following result that expresses the solution of the engines of growth identification problem (7.11) in terms of the reduced TFP growth rate given in (7.12).[3]

Theorem 7.1. *For $1 \leq k \leq n-1$ the k engines of growth $\{i_1^*, i_2^*, \ldots, i_k^*\}$ that solve the problem* (7.11) *give the lowest reduced TFP growth rate, i.e., $g^r_{i_1^*,\ldots,i_k^*} \leq g^r_{i_1,\ldots,i_k}$ for all $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ with $i_s \neq i_r$.*

First, note that in finding the engines of growth from the HEM perspective, given the reduced TFP growth measure in (7.12), performing the traditional procedure of the HEM approach (which includes deleting certain rows and columns of the

[3] Alternatively, given the problem (7.11) we could define the *group TFP growth worth* of sectors $i_1, \ldots, i_k$ as $\omega^g_{i_1,\ldots,i_k} = g - \ell^g_{i_1,\ldots,i_k}$. Hence, in this interpretation, the k engines of growth have the *largest* group TFP growth worth, i.e., $\omega^g_{i_1^*,\ldots,i_k^*} \geq \omega^g_{i_1,\ldots,i_k}$ for all $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$.

matrices $\mathbf{A}$ and $\mathbf{B}$) is not needed at all. Matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{L}$ and $\mathbf{M}$ together with the vectors $\boldsymbol{\pi}$, $\mathbf{p}$ and $\mathbf{x}$ are all given, and only the k identity columns in $\mathbf{E}$ are changed in order to consider all possible combinations of k sectors from the totality of n industries in a search for the k engines of growth. Second, apparently, for fixed k the group of sectors that constitutes the engines of growth depends on the joint sectoral interactions of TFP growth rates, spillover linkages, multiplier effects, and sizes of the gross and net outputs in a complex way as is captured by the reduced TFP growth measure. Let us see this in a simple case when $k = 1$, which implies that one is looking for the *largest engine of growth*. Then $\mathbf{E} = \mathbf{e}_i$, $\mathbf{M}_{kk} = \mathbf{E}'\mathbf{M}\mathbf{E} = \mathbf{e}_i'\mathbf{M}\mathbf{e}_i = m_{ii}$, and $\mathbf{L}_{kk} = l_{ii}$, thus it can be shown that (7.12) reduces to

$$g_i^r = \frac{1}{\mathbf{p}'\mathbf{f} - p_i f_i}\left[\boldsymbol{\lambda}'\mathbf{s} - \frac{\pi_i s_i}{m_{ii}} - \frac{p_i x_i}{m_{ii} l_{ii}}\sum_{k \neq i}(m_{ii}\pi_k - m_{ik}\pi_i) l_{ki}\right], \tag{7.13}$$

which is the *reduced TFP growth* due to extraction of sector i. Then a corollary to Theorem 7.1 is that the single engine of growth, i^*, gives the smallest reduced TFP growth rate, i.e., $g_{i^*}^r \leq g_i^r$ for all $i = 1, \ldots, n$. So it is not only the TFP growth rate of sector i, π_i, that defines it to be the engine of growth, but also its spillover linkage, $s_i = \sum_j m_{ij} p_j x_j$, total (direct and indirect) input self-dependency, l_{ii}, total joint input and capital self-dependency, m_{ii}, and its values of gross and net outputs, $p_i x_i$ and $p_i f_i$, are all important. In particular, (7.13) shows that the engine of growth has a large TFP growth rate and spillover linkage, is less dependent on itself, and, more engaged in the "roundabout" of the production process, hence having higher gross and lower net outputs. But it is the joint relative importance of these factors that defines the engine of growth.

As in the case of key sectors' identification discussed in Chapter 6, it is important to understand that the problem of finding the single engine of growth (i.e., $k = 1$ in (7.11)) is *different* from problem (7.11) with $k > 1$. In other words, k (> 1) sectors, whose extraction results in the smallest reduced TFP growth rates, do *not* necessarily comprise the group of k engines of growth. While the single engine of growth search problem looks for the effect of the extraction of one sector, the more general problem in (7.11) considers the effect of a *simultaneous* extraction of $k \geq 2$ sectors. Hence, the last problem takes into full account all the cross-contributions of the extracted sectors to the nationwide TFP growth that is generated both within and outside the group of sectors. These effects are, of course, differently accounted for when $k = 1$.

Consider two industries that are largely identical with respect to their input and capital linkage patterns (including input and capital stock coefficients' sizes) and that are also similar in terms of their final demands, gross outputs and sources of growth, then their group contribution to the total TFP growth is expected to be less than that of the group consisting of two industries that have quite different patterns of (significant) linkages and TFP growth generation ability. In this case it is said that the first two industries are redundant with respect to each other, hence should not be included *both* in the group with 2 engines of growth. Thus, in general, the k (> 1) sectors, whose individual extraction yields the smallest reduced TFP growth, do *not* comprise the k engines of growth due to the redundancy principle inherent to the majority of real-life input and capital stock networks of interactions of industries.

Computerization is found to have a dramatic impact on growth and structural change by Wolff (2002). In ten Raa and Wolff's (2000) study, the computer and office equipment industry was found to be the largest engine of growth in the US economy for two subperiods of 1967-1977 and 1977-1987, while it was only at the 19-th position in 1958-1967. In the HEM approach discussed above, however, ranking of the individual sectors from the problem of identification of a single engine of growth does not tell us anything about the group of engines of growth. It is the *joint* contribution of sectors to the economy-wide TFP growth generation that makes them engines of growth. In this respect it would be, for example, interesting to find out what is the *minimum* value of k that allows the computers and office machinery industry to be a member of the group of k engines of growth. We plan to do an (extensive) empirical study of the engines of growth determination problem discussed above for several countries, and make a detailed comparison of the results both across countries and over time. This might shed more light on the sectoral analysis of structural change in different countries in terms of industries' contribution to the nation-wide TFP growth rates.

7.3.2 On interregional feedbacks in input-output models

Consider the following interregional input-output model with p regions:

$$\begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^p \end{bmatrix} = \begin{bmatrix} \mathbf{L}^{11} & \mathbf{L}^{12} & \cdots & \mathbf{L}^{1p} \\ \mathbf{L}^{21} & \mathbf{L}^{22} & \cdots & \mathbf{L}^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{L}^{p1} & \mathbf{L}^{p2} & \cdots & \mathbf{L}^{pp} \end{bmatrix} \begin{bmatrix} \mathbf{f}^1 \\ \mathbf{f}^2 \\ \vdots \\ \mathbf{f}^p \end{bmatrix}, \tag{7.14}$$

where

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}^{11} & \mathbf{L}^{12} & \cdots & \mathbf{L}^{1p} \\ \mathbf{L}^{21} & \mathbf{L}^{22} & \cdots & \mathbf{L}^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{L}^{p1} & \mathbf{L}^{p2} & \cdots & \mathbf{L}^{pp} \end{bmatrix} = \begin{bmatrix} \mathbf{I} - \mathbf{A}^{11} & -\mathbf{A}^{12} & \cdots & -\mathbf{A}^{1p} \\ -\mathbf{A}^{21} & \mathbf{I} - \mathbf{A}^{22} & \cdots & -\mathbf{A}^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{A}^{p1} & -\mathbf{A}^{p2} & \cdots & \mathbf{I} - \mathbf{A}^{pp} \end{bmatrix}^{-1}$$

is the Leontief inverse in an interregional setting, $\mathbf{A}^{rr}$ is the (intra)regional input coefficients matrix for region r $(= 1, \ldots, p)$, $\mathbf{A}^{rs}$ is the matrix of interregional input (trade) coefficients with deliveries from region r to region s $(r \neq s)$, $\mathbf{f}^r$ and $\mathbf{x}^r$ are, respectively, the vectors of changes in final demand and gross output for region r, and $\mathbf{I}$ is the identity matrix with appropriate dimension.

The question is how a change of final demand in region, say, 1 (i.e., $\mathbf{f}^1 > \mathbf{0}$ and $\mathbf{f}^r = \mathbf{0}$ for all $r \neq 1$) affects the outputs in that region and what would be the bias if instead of the interregional framework in (7.14) only a single-region input-output framework of $\mathbf{x}_s^1 = (\mathbf{I} - \mathbf{A}^{11})^{-1}\mathbf{f}^1$ would have been used. That is, how big would be the bias in $\mathbf{x}^1$ if the so-called *interregional feedback effects* were totally ignored. The term "feedback" refers to the fact that an increase in final demand in region 1 causes more demand also for the intermediate inputs from other regions, but the production in these regions is in its turn, in general, dependent on the inputs from region 1 as well. Thus other regions will also demand more intermediate goods from region 1. Notice that the final demand in region 1 may also decrease, in which case the directions of all the above mentioned effects will be reversed. If, on the other hand, some components of the final demand in region 1 increase and others decrease, then obviously the impact of the feedback effects is analytically uncertain.

To tackle the above assigned question, for simplicity, the components of (7.14) are reexpressed in terms of two partitioned matrices as follows:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}^\bullet \end{bmatrix} = \left[\begin{array}{c} \mathbf{x}^1 \\ \hline \mathbf{x}^2 \\ \vdots \\ \mathbf{x}^p \end{array}\right], \quad \mathbf{L} = \begin{bmatrix} \mathbf{L}^{11} & \mathbf{L}^{1\bullet} \\ \mathbf{L}^{\bullet 1} & \mathbf{L}^{\bullet\bullet} \end{bmatrix} = \left[\begin{array}{c|ccc} \mathbf{L}^{11} & \mathbf{L}^{12} & \cdots & \mathbf{L}^{1p} \\ \hline \mathbf{L}^{21} & \mathbf{L}^{22} & \cdots & \mathbf{L}^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{L}^{p1} & \mathbf{L}^{p2} & \cdots & \mathbf{L}^{pp} \end{array}\right],$$

and the vector of changes in final demand is $\mathbf{f}' = \begin{bmatrix} (\mathbf{f}^1)' & \mathbf{0}' \end{bmatrix}$, where $\mathbf{f}^\bullet$ is set to zero since there is a change in the final demand for region 1 only, i.e., $\mathbf{f}^r = \mathbf{0}$ for all $r \neq 1$.

The single-region framework can be rewritten as

$$\begin{bmatrix} \mathbf{x}_s^1 \\ \mathbf{x}_s^\bullet \end{bmatrix} = \begin{bmatrix} \mathbf{I} - \mathbf{A}^{11} & -\mathbf{O} \\ -\mathbf{O} & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{f}^1 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} (\mathbf{I} - \mathbf{A}^{11})^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{f}^1 \\ \mathbf{0} \end{bmatrix}, \quad (7.15)$$

which means that in this case in the entire matrix of regional and trade coefficients $\mathbf{A}$, all elements corresponding to any region other than 1 are set to zero. That is, $\mathbf{A}^{1\bullet} = \mathbf{O}$, $\mathbf{A}^{\bullet 1} = \mathbf{O}$ and $\mathbf{A}^{\bullet\bullet} = \mathbf{O}$, where the null matrix $\mathbf{O}$ in each case is assumed to have the appropriate dimension. This nullification is exactly similar to the (generalized) hypothetical extraction method studied in Chapter 6. Hence, we can readily use Lemma 6.2, wherein the setting is now an interregional framework and $\mathbf{E}' = [\mathbf{O}_{\bullet 1} \ \ \mathbf{I}_\bullet]$, where $\mathbf{I}_\bullet$ is the identity matrix of dimension equal to the total number of industries in all regions except region 1, and $\mathbf{O}_{\bullet 1}$ is the null matrix of row dimension equal to the (row or column) dimension of $\mathbf{I}_\bullet$ and column dimension equal to the number of sectors in region 1.[4] Denoting the Leontief inverse in (7.15) by $\mathbf{L}^{-\{\bullet\}}$, we thus have

$$\begin{bmatrix} \mathbf{x}^1 \\ \mathbf{x}^\bullet \end{bmatrix} - \begin{bmatrix} \mathbf{x}_s^1 \\ \mathbf{x}_s^\bullet \end{bmatrix} = \left[\mathbf{L} - \mathbf{L}^{-\{\bullet\}}\right] \begin{bmatrix} \mathbf{f}^1 \\ \mathbf{0} \end{bmatrix} = \left[\mathbf{LE}(\mathbf{E}'\mathbf{LE})^{-1}\mathbf{E}'\mathbf{L} - \mathbf{EE}'\right] \begin{bmatrix} \mathbf{f}^1 \\ \mathbf{0} \end{bmatrix}$$
$$= \left\{ \begin{bmatrix} \mathbf{L}^{1\bullet}(\mathbf{L}^{\bullet\bullet})^{-1}\mathbf{L}^{\bullet 1} & \mathbf{L}^{1\bullet} \\ \mathbf{L}^{\bullet 1} & \mathbf{L}^{\bullet\bullet} \end{bmatrix} - \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{I} \end{bmatrix} \right\} \begin{bmatrix} \mathbf{f}^1 \\ \mathbf{0} \end{bmatrix}.$$

Since we are interested in the effect on outputs in region 1, the last equation yields

$$\mathbf{x}^1 - \mathbf{x}_s^1 = \mathbf{L}^{1\bullet}(\mathbf{L}^{\bullet\bullet})^{-1}\mathbf{L}^{\bullet 1}\mathbf{f}^1 = \mathbf{L}^{1\bullet}(\mathbf{L}^{\bullet\bullet})^{-1}\mathbf{L}^{\bullet 1}(\mathbf{L}^{11})^{-1}\mathbf{x}^1, \quad (7.16)$$

where we have used the fact that $\mathbf{x}^1 = \mathbf{L}^{11}\mathbf{f}^1$ in (7.14) given that the vectors of the change in final demands of all other regions are zero.

Equation (7.16) gives the bias of ignoring interregional feedbacks at the sectoral level of region 1. A widely used measure of an error when a single-region model is used instead of the full interregional framework is the *overall percentage error* (OPE), first employed by Miller (1969) as "*a summary measure of deviation*" (p. 41) and is defined as $OPE = \boldsymbol{\imath}'(\mathbf{x}^1 - \mathbf{x}_s^1)/\boldsymbol{\imath}'\mathbf{x}^1 \times 100$, where $\boldsymbol{\imath}$ is a summation vector of ones. Define the *norm* of any matrix $\mathbf{M}$ as the largest column sum of the absolute values of its elements, and denote it by $\|\mathbf{M}\|$. Therefore, OPE can alternatively be rewritten in terms of norms as $OPE = \|\mathbf{x}^1 - \mathbf{x}_s^1\| / \|\mathbf{x}^1\| \times 100$. For any two matrices $\mathbf{M}$ and

[4] If all regions have the same number of industries equal to n, then $\mathbf{I}_\bullet$ and $\mathbf{O}_{\bullet 1}$ have dimensions of, respectively, $(p-1)n \times (p-1)n$ and $(p-1)n \times n$.

$\mathbf{N}$ the so-called *submultiplicative* property of the matrix norm holds, i.e., $\|\mathbf{MN}\| \leq \|\mathbf{M}\| \, \|\mathbf{N}\|$. Employing this property in (7.16) gives the following proposition.

Theorem 7.2. *The overall percentage error caused by ignoring interregional feedbacks is bounded above by* $\|\mathbf{L}^{1\bullet}(\mathbf{L}^{\bullet\bullet})^{-1}\mathbf{L}^{\bullet 1}(\mathbf{L}^{11})^{-1}\| \times 100$.

From (7.14)-(7.16) it follows that $\mathbf{L}^{11} = (\mathbf{I} - \mathbf{A}^{11})^{-1} + \mathbf{L}^{1\bullet}(\mathbf{L}^{\bullet\bullet})^{-1}\mathbf{L}^{\bullet 1}$. Further, from the theory of (the inverse of) partitioned matrices one can write $(\mathbf{L}^{11})^{-1} = \mathbf{I} - \mathbf{A}^{11} - \mathbf{A}^{1\bullet}(\mathbf{I} - \mathbf{A}^{\bullet\bullet})^{-1}\mathbf{A}^{\bullet 1}$ (see e.g., Sydsæter et al., 2005, p. 140). Using these two identities we have $\mathbf{L}^{1\bullet}(\mathbf{L}^{\bullet\bullet})^{-1}\mathbf{L}^{\bullet 1}(\mathbf{L}^{11})^{-1} = \left[\mathbf{L}^{11} - (\mathbf{I} - \mathbf{A}^{11})^{-1}\right](\mathbf{L}^{11})^{-1} = (\mathbf{I} - \mathbf{A}^{11})^{-1}\mathbf{A}^{1\bullet}(\mathbf{I} - \mathbf{A}^{\bullet\bullet})^{-1}\mathbf{A}^{\bullet 1}$. Thus, the upper bound in Theorem 7.2 can also be rewritten as $\|\mathbf{L}^{1\bullet}(\mathbf{L}^{\bullet\bullet})^{-1}\mathbf{L}^{\bullet 1}(\mathbf{L}^{11})^{-1}\| \times 100 = \|(\mathbf{I} - \mathbf{A}^{11})^{-1}\mathbf{A}^{1\bullet}(\mathbf{I} - \mathbf{A}^{\bullet\bullet})^{-1}\mathbf{A}^{\bullet 1}\| \times 100$. The right-hand side of the last expression (without number 100) is exactly what Guccione et al. (1988) call the *least upper bound* (LUB) of the *OPE*.[5] Thus, in Theorem 7.2 we gave an alternative expression of the LUB in terms of the elements of the Leontief inverse. Note that there is *no* need to use more than a two "region" partition in the analysis of the upper bound for the *OPE*. This has been tried, for example, in Miller (1986) for three-region case, who then not surprisingly noticed that "[t]he algebra is considerably more complex" (p. 292), but more importantly because applying matrix properties to such partitioning might very well result in a looser bound that cannot be the least upper bound.

We should mention that the *OPE* for any other factor than gross output can be easily accommodated in this framework. For this, similar to the discussions in Chapter 6, one should consider also the direct coefficients of the factor of interest (e.g., employment, CO_2 emissions, etc.). In that case the expression for LUB within the norm has to be multiplied by the diagonal matrix of the factor direct coefficients.

Of course, given the ongoing globalization, countries are becoming more and more interdependent not only via trade of final goods, but also through trade of intermediate goods that has been risen steadily over the last several decades. Therefore, one might expect that the error of ignoring interregional feedbacks is much larger now than some 40-50 years ago. This trend of globalization is reflected by more positive and increasing elements in the interregional input coefficient matrices. As a consequence the LUB not surprisingly increases. However, certainly this differs from region to region (or country to country) depending on the self-sufficiencies of the regions. So, the question of how big is nowadays the bias caused by using a single-region framework instead of the multiregional setting is an em-

[5] See also Gillen and Guccione (1980) and Miller (1986) that use a *looser* upper bound.

pirical issue. In the near future, we plan to quantify this bias in the the empirical application part of this section.

7.3.3 Algorithmic considerations of the group intercentrality and group worth measures

In Chapters 5 and 6 we considered the problems of finding the key groups of, respectively, players in networks of social interactions and sectors in an economy. We have also briefly mentioned the complexity issue of finding the exact solutions to these problems for a large number of players/industries and a rather large size of the groups. This is because in order to find the key group of size k, one needs to consider all possible combinations of k players/sectors out of n players/sectors. The number of combination is $C_k^n = n!/(k!(n-k)!)$, which increases exponentially in k and n. For example, in Table 6.3 we have searched for the key groups of size 1 to 4 from a total of 136 sectors, which required to compute the group factor worths of, respectively, 136, 9.180, 410.040, and 13.633.830 different groups. This example clearly demonstrates the problem of the computational complexity inherent to the above mentioned problems. Hence, the cases of searching key group(s) of reasonable size among very large number of groups becomes potentially intractable from a computing point of view. Therefore, the question arises whether for a large n and a rather large k one can find the exact solutions of the key group problems in reasonable time. It turns out that the answer to this question is negative because the posed problems are in the class of the so-called *NP-hard* problems from a combinatorial perspective. NP-hardness implies that there is no possible sophisticated algorithm that will return the exact solution for large n and k in our case. "[N]early all computer scientists ... believe that there is no such algorithm for solving any NP-hard problem. A simple reason for this is that, after decades of continuous search, no one has found efficient algorithm for solving any NP-hard problem" (Ballester et al., 2009, footnote 15). In what follows we first prove that the discussed key group problems are indeed NP-hard problems, and then consider the possible efficient approximate solutions to these problems once the computing search becomes intractable.

Let $N = \{1, 2, \ldots, n\}$ and $z : 2^N \to \mathbb{R}$ be a set function. Nemhauser et al. (1978) considered the following problem:

$$\max_{S \subseteq N} \{z(S) : |S| \leq k, \;\; z(S) \text{ submodular}\}, \tag{7.17}$$

where $|S|$ is the cardinality (i.e., the number of players/sectors) in the set S, while submodularity of a set function is defined as follows.

Definition 7.1. *Given a finite set N, a real-valued function z on the set of subsets of N is called submodular if $z(A) + z(B) \geq z(A \cup B) + z(A \cap B)$ for all $A, B \subseteq N$.*

Without loss of generality z is normalized such that $z(\varnothing) = 0$. We consider non-decreasing set functions in the sense that $z(S) \leq z(T)$ for all $S \subseteq T \subseteq N$. Let us denote the individual contribution by $\rho_i(S) = z(S \cup \{i\}) - z(S)$, which represents the incremental value of adding player/sector i to the set S. Proposition 2.1 in Nemhauser et al. (1978) establishes that an equivalent statement to Definition 7.1 that defines a submodular set function is $\rho_i(S) \geq \rho_i(T)$ for all $S \subseteq T \subseteq N$ and all $i \in N \setminus T$.

From Theorem 5.1 and Theorem 6.3 it follows that the key group problems within the social network and input-output settings are equivalent to the maximization of, respectively, group intercentrality and group factor worth measures. Using the last definition of a submodular function in terms of the individual contributions, in Appendix 7.A we establish the following result.

Lemma 7.1. *The measures of group intercentrality $c_S(\mathbf{g}, a)$ and factor worth $\omega_S^{\pi}(\mathbf{A}, \mathbf{f}, \boldsymbol{\pi})$ are submodular set functions.*

The problem of maximizing a submodular function is NP-hard, in general. Therefore, Lemma 7.1 implies that the computational complexity becomes large when the number of players/sectors n and the group size k are large in the key group problems. Therefore, in such cases *algorithmic approximations* are used, which require much less time in the computation. Consider an *R-step greedy algorithm (heuristic)* that sequentially eliminates the sets of R players/sectors with the highest group intercentrality/worth. Formally, suppose that $k = qR - p$, where q is a positive *integer* and $0 \leq p \leq R - 1$. The R-step greedy heuristic for a set function z works are follows.

Initialization: Let $S^0 = \varnothing$, $S^t = \cup_{i=1}^{t} I^i$, and set $t = 1$.
Iterations: For $t = 1, \ldots, q-1$ select $I^t \subseteq N \setminus S^{t-1}$ with $|I^t| = R$ such that $\zeta_{t-1} = z(S^t) - z(S^{t-1})$ is maximized.
Final step: Choose $I^* \subseteq N \setminus S^{q-1}$ with $|I^*| = R - p$ so as to maximize $z(S^{q-1} \cup I^*) - z(S^{q-1})$.

Consider, for example, the case when we want to find the key group of size $k = 30$ from total $n = 1000$ players using the R-step greedy algorithm, and thus choose

$R = 4$ and $q = 8$. This means that we would like to find out the approximation of the exact key group of size $k = 30$ in $q = 8$ computing steps (iterations). The procedure first sequentially at seven $(q - 1)$ steps eliminates 4 players as a key group (which makes $4 \times 7 = 28$ players). The two remaining members of the key group (i.e., $R - p = k - (q - 1)R = 30 - 28 = 2$, hence $p = 2$) are found in the final stage of the 4-step greedy algorithm. Note that if k is a multiple of R, then $p = 0$.

Let us denote the value of an R-step greedy solution by $z(G^R)$, where the approximate solution set is $G^R = S^{q-1} \cup I^*$, and the exact solution of (7.17) is given by $z(S^*)$. Then, provided the normalization $z(\varnothing) = 0$, the following result is proved in Theorem 4.3 in Nemhauser et al. (1978, pp. 282-283).

Theorem 7.3. *Suppose z is nondecreasing and the R-step greedy heuristic is applied to problem* (7.17). *If $K = qR - p$, with q a positive integer and integer $p \in [0, R - 1]$, then the upper bound of the error of approximation is*

$$\frac{z(S^*) - z(S^{G^R})}{z(S^*)} \leq \left(\frac{q - \lambda}{q}\right)\left(\frac{q-1}{q}\right)^{q-1},$$

where $\lambda = (R - p)/R$.

The bound in Theorem 7.3 for $q > 1$ can be rewritten as (using $K = qR - p$)

$$\left(\frac{q - \lambda}{q}\right)\left(\frac{q-1}{q}\right)^{q-1} = \left(1 + \frac{p}{R(q-1)}\right)\left(\frac{q-1}{q}\right)^{q} < \left(1 + \frac{p}{R(q-1)}\right)\frac{1}{e},$$

where $e \approx 2.718$ is the base of the natural logarithm. The last inequality follows since q is finite and $1/e = \lim_{q\to\infty}(1 - 1/q)^q$. If $p = 0$, then the bound in Theorem 7.3 boils down to $[(q - 1)/q]^q < 1/e \approx 0.3679$. That is, with $p = 0$ the maximum possible error when the R-step greedy algorithm is used to approximate the solution of (7.17) is 36.79%.

Note that if $p = 0$ and $q = 1/R$, then the R-step heuristic is a simple greedy algorithm that selects (eliminates) only one member in each iteration. This is exactly the sequential key player/sector problem that we have discussed in Chapter 5 and Chapter 6. Note that in the input-output setting, we have already shown in Section 6.2.3 that the key group problem is not equivalent to the sequential key sector problem. Since both the group intercentrality and group factor worth are nondecreasing and submodular functions, the result of Theorem 7.3 can be readily used if one wants to approximate the solutions of the key group problems given in (5.6) and (6.4). In particular, it shows that the worst approximation through the sequential key player/sector problem is less than 36.79%. When instead the groups of size R are sequentially selected (i.e., the R-step heuristic with $R > 1$ is used) for the approximation, the same upper bound holds for $p = 0$ (see above), while the error might be larger than 36.79% whenever $p > 0$. These bounds admittedly are very high. Ballester et al. (2009) provide some numerical simulations for 100 different random networks with $n = 10$ and $n = 15$, where they found small approximation errors of at most 1.7% obtained by using a simple greedy algorithm (i.e., the sequential key player problem) in addressing the key group problem in the social network setting.[6] Whether it holds in general for large-sized networks, and for large input-output datasets is a matter that needs deeper investigation.

[6] We should, however, note that the random networks always tend to be more "symmetric", thus they are, in general, different from the real-life networks.

7.A Proofs

Proof of Theorem 7.1. We already know from Lemma 6.2 in Chapter 6 that $\mathbf{L} - \mathbf{L}^{\{i_1,\dots,i_k\}} = \mathbf{L}\mathbf{E}\mathbf{L}_{kk}^{-1}\mathbf{E}'\mathbf{L} - \mathbf{E}\mathbf{E}'$, where $\mathbf{L}_{kk}^{-1} = (\mathbf{E}'\mathbf{L}\mathbf{E})^{-1}$. Note that the last identity holds also if we use the matrix $\mathbf{M}$ instead of $\mathbf{L}$. We further have $\mathbf{f}^{-\{i_1,\dots,i_k\}} = \mathbf{f} - \mathbf{E}\mathbf{E}'\mathbf{f}$, and $\hat{\mathbf{p}}\mathbf{E}\mathbf{E}' = \mathbf{E}\mathbf{E}'\hat{\mathbf{p}}$ since $\mathbf{E}\mathbf{E}'$ is a diagonal matrix. Recalling that the Leontief inverse in the current setting is defined as $\mathbf{L} = (\mathbf{I} - \hat{\mathbf{p}}\mathbf{A}\hat{\mathbf{p}})^{-1}$, we have

$$\begin{aligned}
\hat{\mathbf{p}}\mathbf{x}^{-\{i_1,\dots,i_k\}} &= \mathbf{L}^{-\{i_1,\dots,i_k\}}\hat{\mathbf{p}}\mathbf{f}^{-\{i_1,\dots,i_k\}} = \left(\mathbf{L} - \mathbf{L}\mathbf{E}\mathbf{L}_{kk}^{-1}\mathbf{E}'\mathbf{L} + \mathbf{E}\mathbf{E}'\right)(\hat{\mathbf{p}}\mathbf{f} - \mathbf{E}\mathbf{E}'\hat{\mathbf{p}}\mathbf{f}) \\
&= \hat{\mathbf{p}}\mathbf{x} - \mathbf{L}\mathbf{E}\mathbf{L}_{kk}^{-1}\mathbf{E}'\hat{\mathbf{p}}\mathbf{x} + \mathbf{E}\mathbf{E}'\hat{\mathbf{p}}\mathbf{f} - \mathbf{L}\mathbf{E}\mathbf{E}'\hat{\mathbf{p}}\mathbf{f} + \mathbf{L}\mathbf{E}\mathbf{L}_{kk}^{-1}\mathbf{L}_{kk}\mathbf{E}'\hat{\mathbf{p}}\mathbf{f} - \mathbf{E}\mathbf{E}'\mathbf{E}\mathbf{E}'\hat{\mathbf{p}}\mathbf{f} \\
&= \hat{\mathbf{p}}\mathbf{x} - \mathbf{L}\mathbf{E}\mathbf{L}_{kk}^{-1}\mathbf{E}'\hat{\mathbf{p}}\mathbf{x},
\end{aligned}$$

where the last four terms in the expression after the second equality cancel out since $\mathbf{L}_{kk}^{-1}\mathbf{L}_{kk} = \mathbf{I}$ and $\mathbf{E}\mathbf{E}'\mathbf{E}\mathbf{E}' = \mathbf{E}\mathbf{E}'$. Using $\mathbf{s}^{-\{i_1,\dots,i_k\}} = \mathbf{M}^{-\{i_1,\dots,i_k\}}\hat{\mathbf{p}}\mathbf{x}^{-\{i_1,\dots,i_k\}}$, the above derived expression, and Lemma 6.2, one obtains

$$\begin{aligned}
\boldsymbol{\lambda}'\mathbf{s}^{-\{i_1,\dots,i_k\}} &= \boldsymbol{\lambda}'\left(\mathbf{M} - \mathbf{M}\mathbf{E}\mathbf{M}_{kk}^{-1}\mathbf{E}'\mathbf{M} + \mathbf{E}\mathbf{E}'\right)\left(\mathbf{I} - \mathbf{L}\mathbf{E}\mathbf{L}_{kk}^{-1}\mathbf{E}'\right)\hat{\mathbf{p}}\mathbf{x} \\
&= \boldsymbol{\pi}'\left(\mathbf{I} - \mathbf{E}\mathbf{M}_{kk}^{-1}\mathbf{E}'\mathbf{M}\right)\left(\mathbf{I} - \mathbf{L}\mathbf{E}\mathbf{L}_{kk}^{-1}\mathbf{E}'\right)\hat{\mathbf{p}}\mathbf{x} + \boldsymbol{\lambda}'\mathbf{E}\mathbf{E}'\hat{\mathbf{p}}\mathbf{x} - \boldsymbol{\lambda}'\mathbf{E}\mathbf{L}_{kk}\mathbf{L}_{kk}^{-1}\mathbf{E}'\hat{\mathbf{p}}\mathbf{x} \\
&= \boldsymbol{\pi}'\left(\mathbf{I} - \mathbf{E}\mathbf{M}_{kk}^{-1}\mathbf{E}'\mathbf{M}\right)\left(\mathbf{I} - \mathbf{L}\mathbf{E}\mathbf{L}_{kk}^{-1}\mathbf{E}'\right)\hat{\mathbf{p}}\mathbf{x}.
\end{aligned}$$

Given our objective in (7.11), we are seeking the group of sectors that minimizes

$$\frac{\boldsymbol{\lambda}'\mathbf{s}^{-\{i_1,\dots,i_k\}}}{\mathbf{p}'\mathbf{f}^{-\{i_1,\dots,i_k\}}} = \frac{\boldsymbol{\pi}'\left(\mathbf{I} - \mathbf{E}\mathbf{M}_{kk}^{-1}\mathbf{E}'\mathbf{M}\right)\left(\mathbf{I} - \mathbf{L}\mathbf{E}\mathbf{L}_{kk}^{-1}\mathbf{E}'\right)\hat{\mathbf{p}}\mathbf{x}}{\mathbf{p}'(\mathbf{I} - \mathbf{E}\mathbf{E}')\mathbf{f}}, \tag{7.A.1}$$

which is the definition of the *reduced TFP growth rate* due to extraction of sectors $i_1, \dots, i_k$ in (7.12). ■

Proof of Lemma 7.1. Take $S \subseteq T \subseteq N$ and $i \in N \setminus T$. The strict monotonicity property of the group intercentrality measure discussed in Section 5.2.2 of Chapter 5 immediately implies that $c_{S\cup\{i\}}(\mathbf{g}, a) - c_S(\mathbf{g}, a) = c_i(\mathbf{g}^{-S}, a) \geq c_i(\mathbf{g}^{-T}, a) = c_{T\cup\{i\}}(\mathbf{g}, a) - c_T(\mathbf{g}, a)$, where $\mathbf{g}^{-S}$ denotes the network without all members of the set S. This is exactly the definition of a submodular function in terms of the individual contributions.

Using all the necessary definitions from Chapter 6 and the property of the Leontief inverse matrix, for the industries factor worth $\omega_S^{\pi}(\mathbf{A}, \mathbf{f}, \boldsymbol{\pi})$ we have

$$\omega_{S\cup\{i\}}^{\pi}(\mathbf{A}, \mathbf{f}, \boldsymbol{\pi}) - \omega_S^{\pi}(\mathbf{A}, \mathbf{f}, \boldsymbol{\pi})$$

$$
\begin{aligned}
&= (\boldsymbol{\pi}'\mathbf{x} - \boldsymbol{\pi}'\mathbf{x}^{-\{S\cup\{i\}\}}) - (\boldsymbol{\pi}'\mathbf{x} - \boldsymbol{\pi}'\mathbf{x}^{-S}) \\
&= \boldsymbol{\pi}'\mathbf{x}^{-S} - \boldsymbol{\pi}'\mathbf{x}^{-\{S\cup\{i\}\}} = \omega_i^{\pi}(\mathbf{A}^{-S}, \mathbf{f}^{-S}, \boldsymbol{\pi}) \\
&\geq \omega_i^{\pi}(\mathbf{A}^{-T}, \mathbf{f}^{-T}, \boldsymbol{\pi}) = \boldsymbol{\pi}'\mathbf{x}^{-T} - \boldsymbol{\pi}'\mathbf{x}^{-\{T\cup\{i\}\}} \\
&= \omega_{T\cup\{i\}}^{\pi}(\mathbf{A}, \mathbf{f}, \boldsymbol{\pi}) - \omega_T^{\pi}(\mathbf{A}, \mathbf{f}, \boldsymbol{\pi}),
\end{aligned}
$$

which is again the definition of the submodular function. Hence, both the group intercentrality and group factor worth are submodular set functions. ■